

Credentialing Preclinical Pediatric Xenograft Models Using Gene Expression and Tissue Microarray Analysis

Craig C. Whiteford,¹ Sven Bilke,¹ Braden T. Greer,¹ Qingrong Chen,¹ Till A. Braunschweig,⁴ Nicola Cenacchi,¹ Jun S. Wei,¹ Malcolm A. Smith,⁵ Peter Houghton,⁶ Christopher Morton,⁶ C. Patrick Reynolds,⁷ Richard Lock,⁸ Richard Gorlick,⁹ Chand Khanna,² Carol J. Thiele,³ Mikiko Takikita,⁴ Daniel Catchpoole,¹⁰ Stephen M. Hewitt,⁴ and Javed Khan¹

¹Oncogenomics Section, ²Comparative Oncology Program, and ³Cell and Molecular Biology Section, Pediatric Oncology Branch; ⁴Tissue Array Research Program, Laboratory of Pathology; and ⁵Cancer Therapy Evaluation Program, National Cancer Institute, NIH, Bethesda, Maryland; ⁶St. Jude Children's Research Hospital Memphis, Memphis, Tennessee; ⁷Developmental Therapeutics Program, USC-CHLA, Institute for Pediatric Clinical Research, Children's Hospital Los Angeles, Los Angeles, California; ⁸Leukaemia Biology Program, Children's Cancer Institute Australia for Medical Research, Sydney, New South Wales, Australia; ⁹Section of Hematology/Oncology, The Children's Hospital at Montefiore, Bronx, New York; and ¹⁰Tumour Bank, The Children's Hospital at Westmead, Westmead, New South Wales, Australia

Abstract

Human tumor xenografts have been used extensively for rapid screening of the efficacy of anticancer drugs for the past 35 years. The selection of appropriate xenograft models for drug testing has been largely empirical and has not incorporated a similarity to the tumor type of origin at the molecular level. This study is the first comprehensive analysis of the transcriptome of a large set of pediatric xenografts, which are currently used for preclinical drug testing. Suitable models representing the tumor type of origin were identified. It was found that the characteristic expression patterns of the primary tumors were maintained in the corresponding xenografts for the majority of samples. Because a prerequisite for developing rationally designed drugs is that the target is expressed at the protein level, we developed tissue arrays from these xenografts and corroborated that high mRNA levels yielded high protein levels for two tested genes. The web database and availability of tissue arrays will allow for the rapid confirmation of the expression of potential targets at both the mRNA and the protein level for molecularly targeted agents. The database will facilitate the identification of tumor markers predictive of response to tested agents as well as the discovery of new molecular targets. [Cancer Res 2007;67(1):32–40]

Introduction

Human tumor xenografts have been extensively used for rapid screening of the efficacy of anticancer drugs in the past 35 years (1, 2). However, controversy exists about the usefulness of these preclinical models in predicting response to therapy because many reagents show high activity in these *in vivo* models yet are inactive in the clinical setting (3). This controversy was further fueled by the findings of the National Cancer Institute (NCI) that after 10 years of extensive screening of compounds through preclinical models, only a moderate predictive value was found for their xenograft models,

and even less concordance was found between *in vitro* testing data and clinical usefulness (1).

For pediatric cancers, preclinical *in vitro* and xenograft model systems have been used for drug screening with some success (3). However, because of the rarity of pediatric cancers compared with adult cancers, there has been little emphasis on developing these models by pharmaceutical companies. Consequently, a substantial proportion of pediatric phase 1 trials is being conducted with limited or no prior testing of the agents in pediatric preclinical models (4). Effective prioritization of new agents for clinical testing using reliable preclinical models is especially important in pediatric oncology drug development because of the limited number of children with specific cancer types. Cancer remains the leading cause of disease-related mortality in children >12 months of age, with >2,200 children dying of cancer in the United States alone each year. Future progress in identifying more effective treatments for these children will depend on using reliable preclinical data to select truly active agents for clinical evaluation from among the much larger universe of agents that could be studied.

Several controllable factors contribute to the reliability of xenograft models in predicting *in vivo* drug activity. However, some factors are inherent to these model systems; for example, differences in pharmacokinetic behavior of a drug in mice and humans may render much higher doses of an agent tolerable in mice, leading to a false prediction of clinical activity in humans. Fortunately, these pharmacokinetic differences can be considered and accounted for when interpreting results from these models (3). The use of individual xenograft models rather than a panel of such models may reduce the predictive value because single models cannot capture the inherent variability of the corresponding cancer (5). In addition, certain xenograft models may be poor representations of their purported tumor type of origin. This discordance between the clinical and preclinical entities may go unrecognized because of inadequate biological characterization of the xenograft models. Optimal use of xenograft models for drug testing requires use of panels of xenograft models that closely mimic the biological characteristics of their respective primary tumors and requires consideration of pharmacokinetic differences of tested agents in the human and mice. This report contributes to the optimal use of childhood cancer xenograft models by the molecular characterization of panels of xenograft lines representing many of the more common cancers that occur in children. With this focus on a large panel of models of several pediatric cancers, this study differs from earlier investigations. A panel of 85 xenografts as models of adult

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

C.C. Whiteford and S. Bilke contributed equally to this work.

Requests for reprints: Javed Khan, Oncogenomics Section, Pediatric Oncology Branch, National Cancer Institute, NIH, Room 134E, 8717 Grovemont Circle, Bethesda, MD 20892-4605. Phone: 301-435-2937; Fax: 301-480-0341; E-mail: khanjav@mail.nih.gov.
©2007 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-06-0610

cancers was analyzed by Zembutsu et al. (6). Other investigations typically focused on single adult cancer types, such as prostate cancer (7) or ovarian carcinomas (8). In most cases, a relatively small number of samples were used. For example the work of Mintz et al. (9) on pediatric osteosarcomas used only three xenografts for the verification of expression profiles observed in primary cancer tissues. An analysis of the protein transcription of selected markers was done by Fichtner et al. (10) but this study did not include a genome-wide analysis of expression levels.

A 2001 meeting organized by the NCI and the Children's Oncology Group identified the need for a systematic approach to pediatric preclinical testing to allow the identification of preclinical models that can be used to reliably inform clinical prioritization decisions (11). An early step in implementing the recommendations of the meeting was the Pediatric Oncology Preclinical Protein-Tissue Array Project (POPP-TAP), a collaborative effort between the NCI and the Children's Oncology Group. Xenografts of pediatric tumors were solicited for the POPP-TAP project and a total of 75 high-quality xenografts representing eight tumor types were collected. The majority of these xenografts will be used to screen agents for anticancer activity (11). Objectives of POPP-TAP included developing xenograft tissue microarrays (XMA) for protein expression of a panel of pediatric xenografts and also determining the gene expression profiles of these preclinical models. This study contributes a molecular characterization of a large panel of pediatric xenograft models and determines the extent of their similarity to a set of corresponding primary tumors. In the course of this study, a few xenografts were identified that were not good representations of their primary tumors (i.e., their mRNA profile did not capture the characteristic RNA signature typical for the primary tumors), and these lines have been excluded from further drug testing. This study also shows the use of the xenograft transcriptomic maps along with that of XMA for the discovery of potential new molecular targets applicable to specific childhood cancers.

Materials and Methods

Xenograft, primary tumor, and cell line samples. Samples were acquired through the Pediatric Preclinical Testing Program (PPTP) established by the NCI. An open solicitation for xenografts was made in the journal *Cancer Research* in November 2002. From this solicitation, 95 samples were received for microarray analysis and construction of XMA. All samples received required appropriate Institutional Review Board and Material Transfer Agreement approval from the donating institution. From these 95 tumors, we chose tumor types that had three or more representative xenografts with high-quality RNA. Seventy-five tumors met these criteria and were used for further analysis (Table 1; additional data of the xenograft production are described in Supplementary Table S1).

RNA extraction, amplification, and labeling of cDNA. Total RNA was extracted from the tumors according to the published protocols (12). Agilent BioAnalyzer 2100 (Agilent, Palo Alto, CA) was used to assess the integrity of the total RNAs extracted from all of the samples. Total RNA from seven human cancer cell lines (CHP212, RD, HeLa, A204, K562, RDES, and CA46) was pooled in equal portions to constitute a reference RNA, which was used in all of the cDNA microarray experiments (13). RNA was subjected to one round mRNA amplification using a modified Eberwine RNA amplification procedure (14). Next, an indirect fluorescent-labeling method was used to label cDNA as described by Hegde et al. (15).

Fabrication of cDNA microarrays, hybridization, image acquisition, and analysis. Sequence-verified cDNA libraries were purchased from Research Genetics (Huntsville, AL), and a total of 42,578 cDNA clones,

Table 1. The number of samples for the various primary tumors analyzed in this study

Tumor type	Samples analyzed		
	Xenograft	Primary tumor	Cell lines
Neuroblastoma	20	30	12
Rhabdomyosarcoma	10	21	3
Ewing's	9	19	3
Acute lymphoblastic leukemia	10	—	—
Wilms' Tumor	8	—	—
Osteosarcoma	8	—	—
Medulloblastoma	6	—	—
Ependymoma	4	—	—
Total	75	70	18

representing 13,606 unique genes and 12,327 expressed sequence tags, were printed on microarrays using a BioRobotics MicroGrid II spotter (Harvard Bioscience, Holliston, MA). Fabrication, hybridization, and washing of microarrays were done as described by Hegde et al. (15). Images were acquired by an Agilent DNA microarray scanner (Agilent) and analyzed using the Microarray Suite program as described (16), coded in IPLab (Scanalytics, Fairfax, VA).

Data normalization, filtering, and hierarchical clustering. Gene expression ratios between tumor RNA and reference RNA on each microarray were normalized using a pin-based normalization method modified from Chen et al. (13, 17). To include only high-quality data in the analysis, the quality of each individual cDNA spot was calculated according to Chen et al. (17). Next, spots with an average quality across all of the samples <0.95 were excluded from all of the analyses. There were 38,789 clones that passed this quality filter. All quality-filtered clones (38,789 clones representing 17,349 unique UniGene clusters) were then subjected to hierarchical clustering using a Euclidean distance metric with average linkage (18). Hierarchical clustering was done using the modified Eisen program, Gene Cluster 3.0, and Java TreeView software.¹¹ The entire data set for all 42,421 cDNA clones was released through our Web site.¹² This database allows investigators to make simple queries of the data to extract gene expression profiles based on IMAGE Clone ID, Gene ID (formerly LocusLink), Gene Ontology Terms, Gene Ontology ID, Gene Symbol, UniGene ID, Clone Title, Cytoband, and Chromosome.

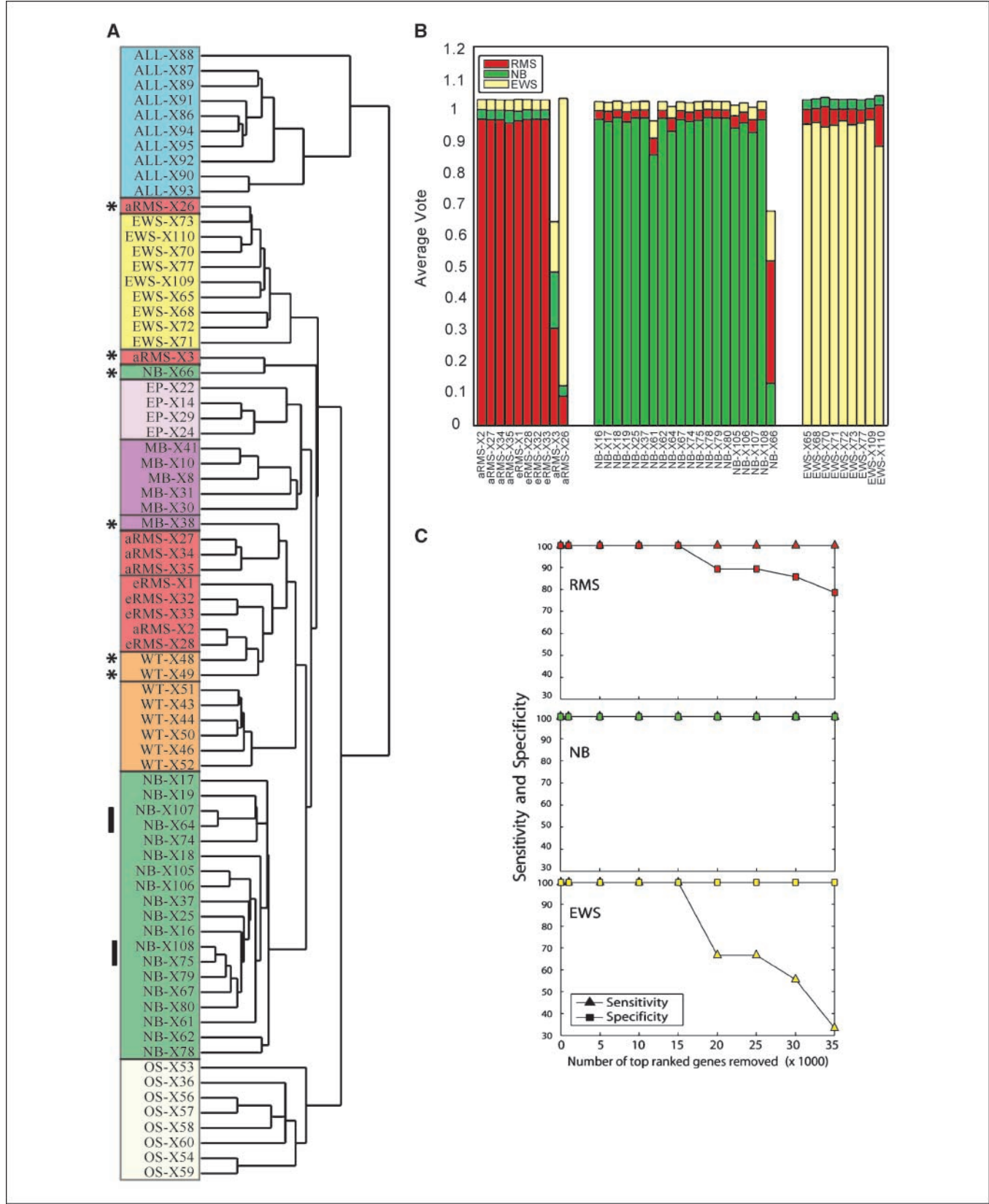
Artificial neural networks and clone-cutter artificial neural network. Feed-forward resilient back-propagation multilayer perceptron artificial neural networks (ANN; coded in Matlab, The Mathworks, Natick, MA) with three layers were used: an input layer of the top 10 principal components of the data; a hidden layer with five nodes; and an output layer generating a committee vote for each of the three input classes. A 4-fold cross-validation scheme with 250 repetitions was used to create 250 "votes" for each sample for each of the three classes (e.g., 0, 0, and 1 or 0.2, 0.8, and 0.3). An average of these ANN committee votes was used to classify samples, and a sample was classified based on the maximum vote it received from the three classes (13, 19). For ANN clone removal analysis, quality-filtered clones were ranked by determining the sensitivity of prediction of the training samples with respect to a change in the gene expression level of each clone. Then, increasing numbers of the top-ranking clones (i.e., the top-ranking 1,000, 5,000, 10,000, 15,000, 20,000, 25,000, 30,000, and 35,000) were cut or removed from both the training and the testing sample data sets, and the ANNs were retrained with the reduced gene sets. The sensitivities and specificities for each of the shaved gene sets were calculated.

¹¹ <http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv>.

¹² <http://home.ccr.cancer.gov/oncology/oncogenomics/>.

The ANN rankings based on training on primary tumor or xenograft expression patterns were compared using Spearman's rank-order correlation, r . The intrinsic statistical variability was estimated by randomly splitting the xenograft data set (after removing the misclassified samples:

aRMS-X3, eRMS-X26, and NB-X66) into two data sets with the relative number of tumor types kept constant and by calculating the gene ranking for each of these. The correlation r was estimated from the coefficients calculated by comparing primary with the xenograft gene ranking and



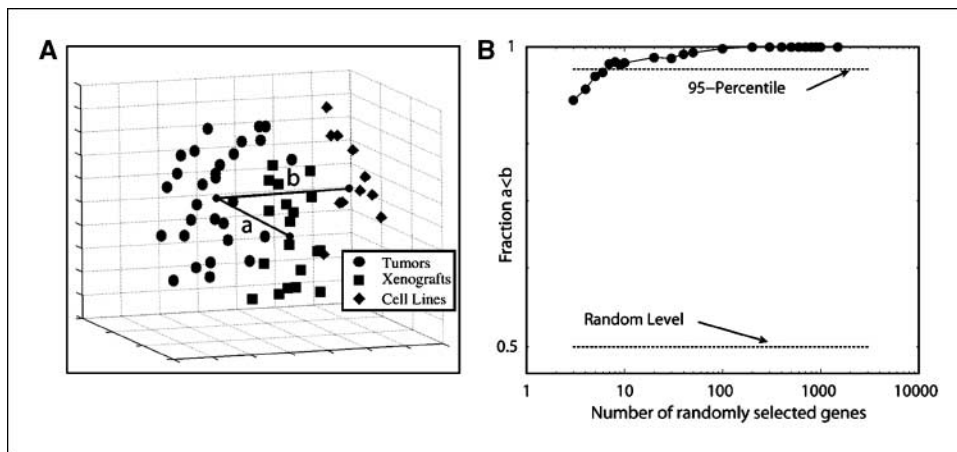


Figure 2. A, multidimensional scaling: multidimensional scaling plot of the 61 neuroblastoma samples (35). Each sphere represents one sample. ●, primary tumor samples ($n = 30$); ■, xenografts ($n = 19$); ♦, cell lines ($n = 12$). Multidimensional scaling is a method to visualize high-dimensional data (here, ~38,789 expression points) in lower dimensions (here, three dimensions), keeping the distance between samples as unchanged as possible. It is used here solely for visualization purposes. The findings reported in the main text are based on the numerical analysis of the distances in the 38,789-dimensional space. B, similarity scaling: the fraction of cases, in which the expression pattern of N randomly selected genes was more similar in xenograft models rather than cell lines compared with primary tumors. For each size, 1,000 iterations with different random genes were probed; the Euclidean metric was used for the comparison (see Materials and Methods).

xenograft with xenograft gene rankings, respectively. The P value indicating nonrandomness of an observed correlation coefficient r was estimated using Student's distribution with

$$t = \sqrt{\frac{N-2}{1-r^2}}$$

and $N - 2$ degrees of freedom (20).

Numerical methods for similarity metric between primary tumor, xenograft models, and cell lines. To measure the similarity between different samples (e1 and e2), the Euclidean metric

$$d_{a,b} = \sqrt{\sum_i^N (a_i - b_i)^2}$$

was used, where the sum runs over N genes. With this metric, a smaller distance indicates that two samples are more similar. The distance between two sets A and B of genes was defined as

$$D(A, B) = \sqrt{\sum_{a \in A, b \in B} (d_{a,b})^2}$$

When comparing a set to itself ($A = B$), this distance measures the spread of the individual samples within the set. When comparing two sets, xenografts X and cell lines C , to a third set of samples, primary tumors P , the set X was said to be more similar to P when $D(X, P) < D(C, P)$. To test how dependent such a result was on the specific choice of genes used in the comparison, the numerical experiment was repeated for different randomly selected subsets of genes. For various subset sizes, the fraction of cases out of 1,000 repeats for which $D(X, P) < D(C, P)$ was counted. The same type of experiment was done using the Pearson's correlation as a metric. When comparing two sets, the average correlation coefficient (as opposed to the

average squared distance used with the Euclidean metric) was used. The set X was said to be more similar to P than the set C if $D(X, P) < D(C, P)$ with the correlation as the metric.

Identifying cancer-specific gene targets. Differentially expressed genes were first identified by doing a t test analysis to identify genes whose mean ratio was significantly higher in xenograft compared with normal tissues ($n = 76$ samples). Clones were selected using the criteria that the Bonferroni adjusted P values was < 0.01 ($n = 14,489$). Next, the list was further filtered by requiring that the median ratio in xenografts be five times greater than the median ratio of normal tissues ($n = 248$). Any clone that belonged to either zero or multiple UniGene clusters or expressed sequence tags was then removed (remaining, $n = 157$). Finally, redundant clones in UniGene cluster represented by multiple clones were removed by removing all but the highest ranked clone ($n = 120$).

XMA construction. Frozen xenograft samples were defrosted to room temperature > 5 min, sectioned to appropriate thickness (2–3 mm), placed in processing cassettes, and fixed in 70% ethanol at 4°C (21). Ethanol was chosen as a fixative instead of formalin because it is useful for downstream proteomic analysis as planned (22) and offers many advantages as follows. Ethanol is a non-cross-linking fixative that can be used to replace formalin where recovery of native proteins and intact nucleic acids is desired. In addition, immunohistochemistry on ethanol-fixed tissue requires less or no antigen retrieval compared with formalin-fixed tissues. In addition, a greater fraction of antibodies doing well in Western blot can be used when comparing ethanol-fixed to formalin-fixed tissue. Our choice thus makes the XMA particularly suitable for testing antibodies not commonly used in immunohistochemistry (21). However, we will also offer a XMA built with formalin-fixed tissues, which allows the use of conventional diagnostic antibodies.

After ethanol fixation for 48 h, the specimens were processed and infiltrated with paraffin and subsequently embedded for sectioning. H&E sections were made of each xenograft and reviewed to select appropriate areas (zones without necrosis) of the xenograft for arraying. The XMA was

Figure 1. A, hierarchical clustering of xenografts. All quality clones ($N = 38,789$) of the preclinical pediatric xenograft models were subjected to hierarchical clustering with average linkage using Pearson's correlation coefficient as the metric. Distinct colors were used for each tumor type to enhance readability. *, xenografts that did not cluster with the majority of the same cancer type. Solid vertical lines, two pairs of samples each derived from a common cell line (SK-N-AS: NB-X75 and NB-X107; SMS-KCNR: NB-X75 and NB-X108) but obtained from different laboratories (see Supplementary Table S1) and derived from the s.c. (X75 and X107) or orthotopic (intra-adrenal; X108 and X107) route. B, ANN average committee votes from a feed-forward resilient back-propagation multilayer perceptron ANN with three layers: an input layer of the top 10 principal components of the data; a hidden layer with five nodes; and an output layer generating a committee vote for each of the three input classes. C, clone cutter: sensitivity and specificity with increasing number of the top-ranking clones removed from the training and testing data sets. Quality-filtered clones were ranked by determining the sensitivity of prediction of the training samples with respect to a change in the gene expression level of each clone. Then, after classification using the clone-cutter ANN, the sensitivities (true positives) and specificities (true negatives) of the ANN to predict the xenograft samples were calculated with each successive removal of the top-ranking ANN clones and plotted.

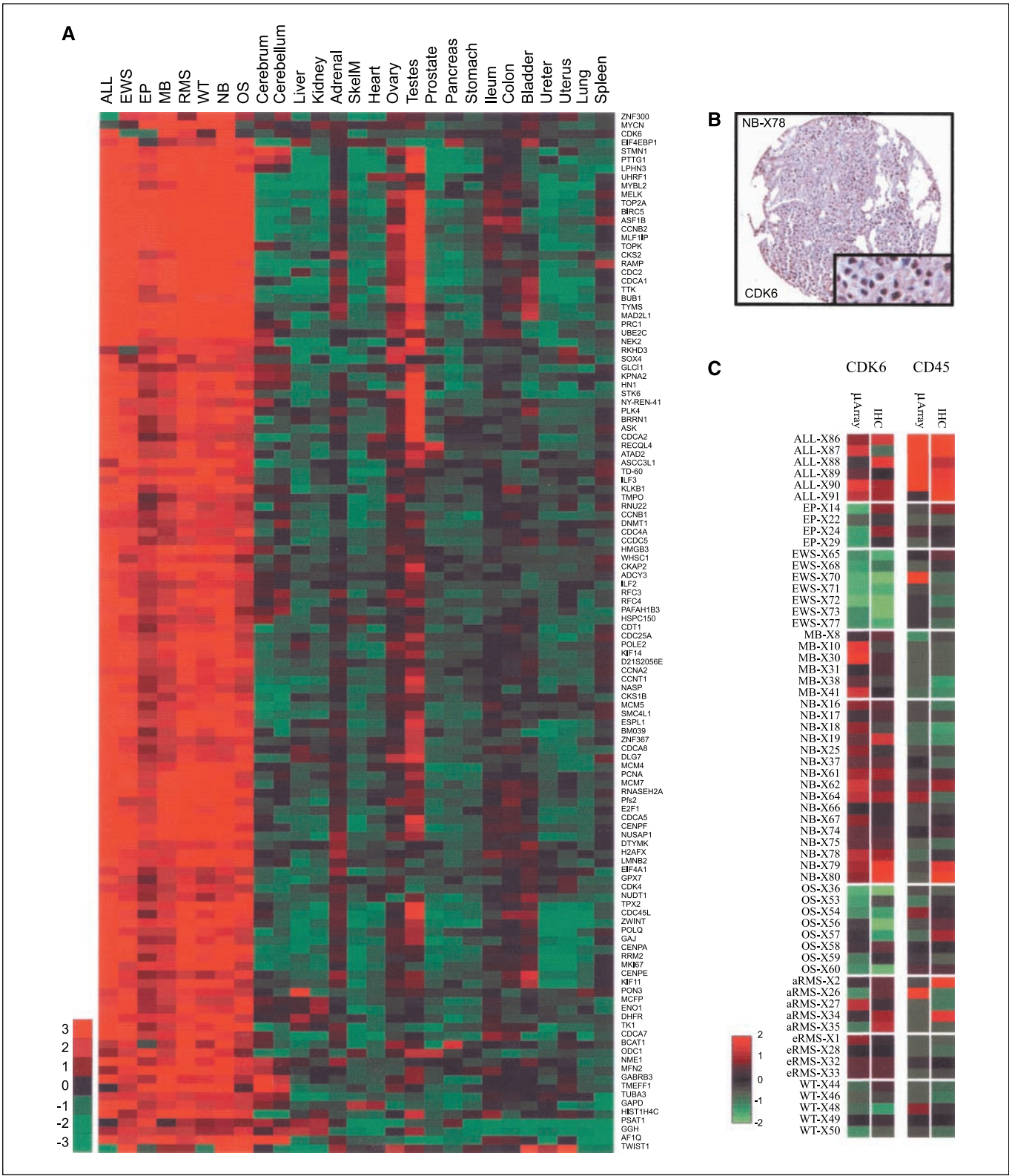


Figure 3. A, identifying cancer-specific gene targets: heatmap of median tissue and tumor values for 120 differentially expressed genes between xenografts and normal tissues. A combination of *t* test (Bonferroni adjusted $P < 0.01$) and fold change (>5) filtering identified 248 clones, which mapped to 157 unique UniGene clusters, out of which 120 were known genes. Many of the genes belong to the functional annotation categories of cell cycle, cell division, DNA metabolism, and other gene ontology annotations that would suggest good therapeutic targets (Supplementary Table S4). B, tissue microarray image: positive CDK6 stain (nuclear) of a xenograft derived from a neuroblastoma cell line (SK-N-DZ). Original magnification, $\times 200$. C, comparison of RNA expression and protein expression from gene expression microarrays and tissue microarrays, respectively. The data for the protein expression is the fraction of total cells that stained positive as measured by the Aperio ScanScope. Both the RNA and the protein expression data were z scored before heatmap visualization. Samples with missing cores on the tissue microarrays were excluded.

constructed as described previously (23), using 1.00-mm needles on a Beecher manual tissue microarrayer MTA-1 (Beecher Instruments, Sun Prairie, WI). The resultant recipient XMA block was sectioned into 5- μ m sections with the aid of an Instrumedics tape sectioning system (St. Louis, MO). Our XMA's are available for investigators to confirm the protein expression levels of their own target(s) of interest. We strongly encourage submission of the images of the immunostains of the XMA to our databases as we have done.

Immunohistochemistry and scoring. Immunohistochemistry was done according to standard protocols as described previously (24). The antibody against CD45 was obtained from DAKO (Carpinteria, CA) and used at titers with incubation times as follows: prediluted, 60 min, room temperature; anti-cyclin-dependent kinase 6 (CDK6) was obtained from Santa Cruz Biotechnology (Santa Cruz, CA) and used at 1:50 titer with an overnight incubation time at 4°C. Preceding treatment of the slides was an antigen retrieval at 95°C for 25 min with antigen retrieval solution (DAKO), endogenous peroxidase blocking, and unspecific binding blocking. All antibodies were detected with the LSAB2 system and 3,3'-diaminobenzidine as the colorizing step (DAKO). Immunostains were reviewed both manually and with the aid of automated image analysis. An Aperio T2 Scanscope (Aperio, Vista, CA) was used to generate high-resolution images of the XMA. These images were quantitatively analyzed (24) with Aperio image analysis software using appropriate algorithms for membranous and nuclear staining. For membranous staining, a ratio of the number of positive (brown) pixels to the sum of all pixels was calculated. For nuclear staining, a ratio of positive (brown) nuclei to the sum of all nuclei was calculated.

Web-based database. We have released the gene expression data from the xenografts, tumor tissues, and cell lines¹² as well as immunohistochemistry images. The web interface offers a broad variety of options for data query, normalization, and visualization. It also offers an option to compare the expression profiles to our expression database of normal human tissues (25).

Results

Hierarchical clustering of xenograft models. To determine the transcriptomic consistency among the panel of pediatric xenograft models, cDNA microarray analysis was done on 75 preclinical pediatric xenograft models, 70 primary tumors samples, and 18 cell lines (Table 1). All quality-filtered clones (38789; ref. 17) were subjected to hierarchical clustering (18). Figure 1A shows that the xenograft models primarily clustered according to their prospective tumor types with the exception of 6 of 75 xenografts analyzed. Of note, in two cases where xenografts were derived from the same original cell lines but were propagated through distinct pathways (s.c. and intra-adrenal in two different laboratories), the corresponding samples clustered closest to each other, suggesting that the global expression profile is not strongly affected by the choice of the site of implantation. The xenografts that did not cluster according to their tumor type were WT-X48, WT-X49, MB-X38, NB-X66, aRMS-X3, and eRMS-X26.

ANN classification of xenograft models. To establish if the xenografts maintained the characteristic expression pattern of the primary tumors, cDNA microarray analysis was done on an additional set of primary tumor tissue consisting of 19 Ewing's tumor, 22 rhabdomyosarcomas, and 30 neuroblastomas (Supplementary Table S2). ANNs were trained with all of the quality-filtered clones (38 and 789) in the primary tumors and then classified or tested the respective xenograft samples (13, 19). In Fig. 1B, the average ANN vote is shown for the xenografts (the test set). This graph shows that the ANN predicted the respective xenograft tumors based on the expression profiles present in the primary tumors with the exception of NB-X66, aRMS-X3, and eRMS-X26. To further show the similarity of the xenografts with the primary

tumors, the same analysis was done with the xenografts as the training set (with the three misclassified samples removed) and the primary tumors as the test samples. The resulting classifier was able to classify all of the primary tumors correctly (Supplementary Table S3).

To determine if the clones used by the classifier extended beyond a particular small subset of clones or if there is a larger set of discriminating clones, another ANN analysis was done, in which the top ANN-ranked clones were sequentially removed. Figure 1C shows that the sensitivity and specificity of the primary tumors to predict respective xenografts remained at 100%, even with the 25,000 most informative clones, almost two thirds of the entire data set, were removed. This analysis (referred to as clone-cutter ANN) showed that many different subsets of genes were equally capable of distinguishing the different tumor types.

Spearman's rank-order correlation of ANN-ranked clones. Our ANN analysis showed that it is possible to develop a broad classifier on the model systems, which could in turn predict primary tumors and vice versa. However, for many applications of the xenograft expression database (e.g., the identification of markers), it is necessary that the "importance" attributed to a gene does not depend on whether it was estimated from xenograft or primary tumor data. The weight of a genes contribution to the classifier, the so-called ANN rank, is frequently used to select potentially biological important genes (13, 19). The ranking of a gene should therefore not differ when xenografts or primary tumors were used to develop the classifier. The degree of similarity of the ranking was determined by calculating Spearman's rank-order correlation between the list of ANN-ranked clones when primary tumors were used to train the ANN and the list obtained when xenografts were used for training. Although the observed correlation $r = 0.67$ ($P < 0.001$) was smaller than a perfect correlation $r = 1$, it was strong. Potential contributors for $r < 1$ are statistical noise and true biological differences. To estimate how much each contributed, the "normal" statistical fluctuation was estimated by calculating the correlation between two lists without systematic biological differences: xenografts were split in two nonoverlapping groups and the rank order was estimated for each group individually. The correlation for these two groups was 0.76 ($P < 0.001$), only moderately higher than $r = 0.67$ for primary tumors/xenografts. This suggests that differences in the lists of the ANN-ranked genes trained on either primary tumors or xenografts are mostly of statistical nature and reflect only weakly systematic differences in gene expression.

Multidimensional scaling and similarity metric between primary tumor, xenograft models, and cell lines. Up to this point, the overall high transcriptional similarity has been shown between the xenografts and their respective primary tumor types. An even more immediate way to measure the similarity of expression profiles is to calculate the distance between expression vectors using some metric. The average Euclidean distance between all pairs of xenografts and primary human tumors was $E_x = 0.622 \pm 0.002$. Obviously, such a pure number is difficult to interpret because it lacks a scale. Comparison to another established model system, cell lines, provided a reference point. For additional 12 neuroblastoma cell lines, the average distance to primary tumors was $E_c = 0.757 \pm 0.004$. The difference between the E_x and E_c , >30 SEs, was highly significant with $P < 0.0001$ (20), indicating that, compared with cell lines, xenograft expression was closer to primary tumors. The experiment was repeated using Pearson's correlation r as the metric. Again, the xenografts were

significantly ($P < 0.0001$; ref. 20) more similar to the primary tumors ($1 - r = 0.43$) than the cell lines ($1 - r = 0.56$). Figure 2A visualizes these results.

The global comparison does not exclude the possibility that only a small, specific set of clones is more similar in the xenograft models, whereas in the remaining transcriptome cell lines could be equally similar or even closer to primary tumors. To address this question, random subsets of clones were chosen and the number of cases where the xenograft expression pattern was more similar to the primary tumors than the cell lines was counted. Subset sizes ranged from 3 clones to 2,000 clones with 1,000 randomly generated subsets for each subset size (Fig. 2B). Even when selecting only three genes, the xenografts were closer to the primary tumors in >80% of the cases. Interestingly, this value increased quickly with the size of the gene subset; the 95% level was achieved with only 10 genes. For sets as small as 100 random genes, the neuroblastoma xenograft models were always found to be more similar to the primary tumors than the cell lines.

Identification of potential therapeutic targets. One application of the xenograft expression database is to identify uniquely expressed genes, potential diagnostic markers, or targets for therapy. The identification of such genes of interest depends on the exact biological question and the method used to extract these genes. For this reason, we have released the entire gene expression data set to enable other researchers to develop their own optimized queries and do simple searches and compare gene, the expression level of that gene, with normal tissues. The data can also be downloaded or queried using a versatile user interface on our Web site.¹² As one possible example of a gene identification, we compared the xenografts with a previously published gene expression database of normal organs. An on-line version of this database is available online¹³ (25). It was found that 120 known genes were up-regulated (Fig. 3A) with many genes involved in cell cycle, cell division, DNA metabolism, and other gene ontology annotations that may be good therapeutic targets (Supplementary Table S4).

One challenge in identifying biomarkers or drug targets based on gene expression analysis is the fact that in some instances the transcripts may not correlate with the protein levels due to post-transcriptional and translational regulation (26, 27). The XMA developed for this study thus enables a further selection of potential markers based on protein expression. As a demonstration, we chose one established clinically useful markers, CD45 (lymphoid malignancies), as well as one of the 120 up-regulated genes, CDK6 (Fig. 3B). This cell cycle protein was chosen as potentially “druggable” as it showed an elevated expression level in the xenograft models for acute lymphoblastic leukemia, medulloblastoma, rhabdomyosarcoma, and neuroblastoma. Automated image analysis of the XMA provided quantitative data of protein expression. A significant ($P < 0.001$; ref. 20) Pearson's correlation r between the protein signal and the mRNA level was observed (Fig. 3C) for all tested genes: $r = 0.42$ (CD45) and $r = 0.55$ (CDK6). The images are also released in our on-line gene expression database.¹²

Discussion

A fundamental assumption in using human tumor xenografts as models for preclinical anticancer drug development is that the

xenografts closely resemble the corresponding primary tumors. Previous studies have analyzed the similarity of xenograft models to primary tumors by comparing specific biological phenotypes of the primary tumor, such as tumorigenicity (28), tumor volume (29), or DNA index (30). Here, we have taken a more systemic approach. Rather than focusing on one specific aspect of tumor biology, we quantified transcriptional similarities on scales ranging from only a few genes to a level of thousands of genes. We and others have shown that such profiles reflect the overall biology of cancers (13, 25, 31).

The first step of our analysis was to do a global survey of our expression data, which also served to ensure internal consistency of the data set. Using hierarchical clustering, we verified that specific tumor types have similar expression patterns to themselves and that they clustered according to their respective tumor type. Hierarchical clustering showed that the majority of the xenografts grouped according to their specific tumor types (with the exception of six xenografts), which also established the internal consistency of our data set.

The second step was to formally validate that the expression profiles of the xenografts reflect those of the corresponding primary tumors. ANNs on three sets of tumors (Ewing's tumor, rhabdomyosarcoma, and neuroblastoma) were used to test if the characteristic patterns discriminating different tumor types in primary tumors were preserved in the xenograft models. The ANN trained with profiles of primary tumors could accurately diagnose the xenograft tumors for the majority of xenograft models. The ANN rank assigned to a specific gene was similar regardless of the samples (xenografts or primary tumors) used to train the ANN. The variations of gene ranks between xenografts and primary tumor-generated classifiers could be mostly explained by statistical uncertainty. The stability of the ranking between model system and primary tumor therefore suggests that the xenograft gene expression database is an effective tool also for marker discovery, particularly in combination with the XMA.

Next, we used the Euclidean and Pearson's distance of expression profiles as the most immediate way to measure profile similarity. The average distance of model systems from primary tumors indicates how well the model represents the primary tumor. Interestingly, the comparison of the results for two clinical model systems in neuroblastoma revealed that xenografts were significantly closer to primary tumors than cell lines were. This suggests that this higher level of similarity on the mRNA level may translate also to a higher level of similarity in the physiologic response to a drug. Of particular importance in this context is the finding that this higher level of similarity holds not only on the systemic scale, the entire transcriptome, but also for smaller, randomly selected sets of genes. Naturally, one can think of a drug affecting the function of particular biological pathways with only a modest number of genes involved. On the scale of pathways (we used 100 genes in our experiments), we observed that xenografts were closer to the primary tumors in all of the 1,000 probed random sets of genes. Even on a “microscale”, 10 genes, this remained true in 95% of the cases. This finding may have implications on the choice of model systems for the testing of drugs, in that the xenograft models might be a better choice especially when testing drugs with unknown targets or multiple targets (i.e., so-called “dirty drugs”). In other words, this analysis for neuroblastoma xenografts indicates that it is highly probable to find a higher level of similarity to primary tumors not only for a particular gene target or targets but also for the genes in the context (i.e., the biological system), in which these targeted genes

¹³ http://ntdtdb.abcc.ncifcrf.gov/cgi_bin/nttissue.pl.

function. Because the transcriptional similarity reflects the overall biology of the cancer (13, 25, 31), it is conceivable that a higher level of transcriptional similarity would be additive to the predictive value of the xenograft model system. Interestingly, the neuroblastoma xenograft samples in this analysis were not direct transplants but rather derived from cell lines. The observed change of their expression profiles from that of cell lines toward that of primary tumor tissues therefore suggests that this shift is induced by the microenvironment emulated by the foreign host organism. Still, the multidimensional scaling in Figure 2A clearly indicates that significant systematic differences in expression levels of primary tumors and xenografts remain. This is partially explained by the fact that the human cDNA array in this study is relatively insensitive to mouse RNA due to differences in the 3' sequences of mouse and human RNA. A separate hybridization of only mouse RNA to our microarray showed low signal intensities (Supplementary Fig. S1). RNA from stroma cells or blood vessels, which are present in both human and xenograft samples, are therefore detected only in the primary samples. An analysis of the differences as well as the pattern of the mouse stroma will be subject of future studies.

Of note, the ANNs trained on tumor samples rejected the very same xenografts (NB-X66, aRMS-X3, and eRMS-X26) that did not cluster with their respective tumor type in the hierarchical clustering, thus emphasizing the internal consistency of our data and the concordance of our analysis. The NB-X66 and aRMS-X3 xenografts neither clustered nor classified with any other xenograft tumors present in our analysis. In the hierarchical clustering, these two xenografts shared a common and isolated branch, suggesting that they share some common features, but not related to their original diagnostic assignment, and thus would not be a good model to test drugs targeted against neuroblastoma or RMS, respectively. Of the remaining misclassified samples (eRMS-X26, WT-X48, WT-X49, and MB-X38), the ANNs classified eRMS-X26 as a Ewing's tumor. Interestingly, eRMS-X26 was initially diagnosed as embryonal rhabdomyosarcoma; however, review by others has reclassified it as a primitive neuroectodermal tumor, which is a member of the Ewing's family of tumors (32). The presence of the EWS-FLI translocation in this xenograft was confirmed by reverse transcription-PCR analysis (data not shown). Therefore, our transcriptomic analysis was able to correctly diagnose on a global scale a xenograft that was initially misdiagnosed. The two Wilms' xenografts (WT-X48 and WT-X49) and the medulloblastoma (MB-X38) clustered with the rhabdomyosarcoma. This is not surprising given the fact that both of these groups of cancers have been reported on occasions to express muscle markers. Indeed, medulloblastomas are a heterogeneous group of cancers and the majority of reported cases in the literature have been biphasic, containing both primitive neuroectodermal and rhabdomyoblastic cells (33, 34). Within our database, all three of these tumors were

found to have the highest expression of insulin-like growth factor II in each of their cancer types at levels comparable with levels in rhabdomyosarcoma (data not shown; see on-line database).¹² This observation leads us to conclude that the heterogeneity of the Wilms' tumor and medulloblastoma xenografts reflects that of the tumor group that they were derived from and is one of the strengths of using a panel of xenografts rather than individual models for drug testing.

The XMA presented in this work will be of particular usefulness for the confirmation of protein expression detectable by immunohistochemistry. The platform offers a high-density approach, which is not feasible with Western blots and additionally offers the benefit of histomorphology. The intention to make the XMA available to other research groups essentially ruled out an array based on frozen tissue. We decided to use ethanol as a fixative, which is "proteomic friendly" (22) and offers many advantages as outlined in Materials and Methods. However, all fixation protocols can be associated with loss of localization of antigens, and non-cross-linking fixatives, such as ethanol, are associated with less subcellular detail; nevertheless, ethanol fixation has become the fixation of choice for nonclinical samples to allow a broader range of investigations to be done on these samples.

In conclusion, we have characterized the gene expression profiles of a large panel of pediatric xenografts and have established that xenografts closely resemble their tumor types of origin even at the level of 10 randomly selected genes. Many of the xenografts that we have shown to resemble their tumor type of origin have been subsequently incorporated into the NCI-sponsored PPTP (11). These tumor panels will be used to systematically evaluate the activity of 10 to 15 new agents yearly that are being considered for clinical evaluation in children with cancer. The transcriptional profiles and XMA's described here will make key contributions to the PPTP. Our web database and tissue arrays (see Materials and Methods) will make possible the rapid confirmation of potential targets at both the mRNA and protein level for available molecularly targeted agents. Finally, our data should facilitate the identification of tumor markers predictive of response to tested agents as well as the discovery of new molecular targets applicable to specific childhood cancers.

Acknowledgments

Received 2/16/2006; revised 9/7/2006; accepted 10/27/2006.

Grant support: This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Drs. Peter Adamson and John Maris from the Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine and the Children's Oncology Group Phase 1 Consortium for fruitful discussion and insightful comments during the planning stages of this study.

References

- Johnson JJ, Decker S, Zaharevitz D, et al. Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br J Cancer* 2001; 84:1424-31.
- Sausville A, Burger AM. Contribution of human tumor xenografts to anticancer drug development. *Cancer Res* 2006;66:3351-4.
- Peterson JK, Houghton PJ. Integrating pharmacology and *in vivo* cancer models in preclinical and clinical drug development. *Eur J Cancer* 2004;40:837-44.
- Houghton PJ, Adamson PC, Blaney S, et al. Testing of new agents in childhood cancer preclinical models: meeting summary. *Clin Cancer Res* 2002;8: 3646-57.
- Voskoglou-Nomikos T, Pater JL, Seymour L. Clinical predictive value of the *in vitro* cell line, human xenograft, and mouse allograft preclinical cancer models. *Clin Cancer Res* 2003;9:4227-39.
- Zembutsu H, Ohnishi Y, Tsunado T, et al. Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res* 2002;62: 518-27.
- Glinsky GV, Krones-Herzig A, Glinskii AB, Gebauer G. Microarray analysis of xenograft-derived cancer cell

- lines representing multiple experimental models of human prostate cancer. *Mol Carcinog* 2003;37:209–21.
8. Bani MR, Nicoletti MI, Alkharouf NW, et al. Gene expression correlating with response to paclitaxel in ovarian carcinoma xenografts. *Mol Cancer Ther* 2004;3:111–21.
 9. Mintz M, Sowers R, Brown KM, et al. An expression signature classifies chemotherapy-resistant pediatric osteosarcoma. *Cancer Res* 2005;65:1748–54.
 10. Fichtner I, Slisow W, Gill J, et al. Anticancer drug response and expression of molecular markers in early-passage xenotransplanted colon carcinomas. *Eur J Cancer* 2004;40:293–307.
 11. Houghton PJ, Gorlick R, Friedman H, et al. Pediatric Preclinical Testing Program (PPTP)—a molecularly characterized panel of childhood cancer models for new agent testing [abstract 4731]. Proceedings of the 96th Annual Meeting of the AACR; 2005. p. 1116–7.
 12. Wei JSKJ. Purification of total RNA from mammalian cells and tissues. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2002.
 13. Wei JS, Greer BT, Westermann F, et al. Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. *Cancer Res* 2004;64:6883–91.
 14. Eberwine J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *Biotechniques* 1996;20:584–91.
 15. Hegde P, Qi R, Abernathy K, et al. A concise guide to cDNA microarray analysis. *Biotechniques* 2000;29: 548–50, 52–4, 56 passim.
 16. Chen Y, Kamat V, Dougherty ER, Bittner ML, Meltzer PS, Trent JM. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 2002;18:1207–15.
 17. Chen YJ, Kodell R, Sistare F, Thompson KL, Morris S, Chen JJ. Normalization methods for analysis of microarray gene-expression data. *J Biopharm Stat* 2003;13:57–74.
 18. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
 19. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7: 673–9.
 20. Press WH, Tutorials SA, Vetterling WT, N Flannery BP. Numerical recipes in C. 2nd Ed. Cambridge University Press; 1992.
 21. Gillespie JW, Best CJ, Bichsel VE, et al. Evaluation of non-formalin tissue fixation for molecular profiling studies. *Am J Pathol* 2002;160:449–57.
 22. Ahram M, Flaig MJ, Gillespie JW, et al. Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. *Proteomics* 2003;3:413–21.
 23. Khanna C, Wan X, Bose S, et al. The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nat Med* 2004;10:182–6.
 24. Tan AR, Yang X, Hewitt SM, et al. Evaluation of biologic end points and pharmacokinetics in patients with metastatic breast cancer after treatment with erlotinib, an epidermal growth factor receptor tyrosine kinase inhibitor. *J Clin Oncol* 2004;22:3080–90.
 25. Son CG, Bilke S, Davis S, et al. Database of mRNA gene expression profiles of multiple human organs. *Genome Res* 2005;15:443–50.
 26. Chen G, Gharib TG, Huang CC, et al. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 2002;1:304–13.
 27. Beyer A, Hollunder J, Nasheuer HP, Wilhelm T. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol Cell Proteomics* 2004;3:1083–92.
 28. Verschraegen CF, Hu W, Du Y, et al. Establishment and characterization of cancer cell cultures and xenografts derived from primary or metastatic Mullerian cancers. *Clin Cancer Res* 2003;9:845–52.
 29. Schmidt KF, Ziu M, Schmidt NO, et al. Volume reconstruction techniques improve the correlation between histological and *in vivo* tumor volume measurements in mouse models of human gliomas. *J Neurooncol* 2004;68:207–15.
 30. Verstijnen CP, Arends JW, Moerkerk P, et al. Culturing and xenografting of primary colorectal carcinoma cells: comparison of *in vitro*, and *in vivo* model and primary tumor. *Anticancer Res* 1988;8:1193–200.
 31. Khanna C, Khan J, Nguyen P, et al. Metastasis-associated differences in gene expression in a murine model of osteosarcoma. *Cancer Res* 2001;61:3750–9.
 32. Morton CL, Potter PM. Rhabdomyosarcoma-specific expression of the herpes simplex virus thymidine kinase gene confers sensitivity to ganciclovir. *J Pharmacol Exp Ther* 1998;286:1066–73.
 33. Misugi K, Liss L. Medulloblastoma with cross-striated muscle. A fine structural study. *Cancer* 1970; 25:1279–85.
 34. Helton KJ, Fouladi M, Boop FA, et al. Medulloblastoma: a radiographic and clinicopathologic analysis of six cases and review of the literature. *Cancer* 2004;101: 1445–54.
 35. Khan J, Saal LH, Bittner ML, Chen Y, Trent JM, Meltzer PS. Expression profiling in cancer using cDNA microarrays. *Electrophoresis* 1999;20:223–9.